Journal of the Korean Institute
of Industrial Engineers
Vol.22, No.1, March, 1996

An Analysis of the Behavior of Correlated Arrival Queues

Hur, Sun*

Abstract

In this research, we concentrate on the effects of dependencies in arrival processes on queueing measures. In particular, we use a specific form of arrival process which has the advantage of allowing us to change dependency properties without at the same time changing one dimensional distributional conditions.

It is shown that the mean queue length can be made arbitrarily large with the same interarrival distributions and the same service time distributions with fixed smaller than one traffic intensity.

1 Introduction

Nonrenewal[1] arrival processes are common, for example, in manufacturing systems, where simplifying independence assumptions can lead to very poor estimates of performance measures. Little is known regarding the queues with a nonrenewal input, because of loss in analytical tractability.

Two prototypes of non-renewal arrival processes are double stochastic Poisson (DSP) and Markov renewal (MR) processes. Some basic results on MR arrival queues are contained in Çinlar (1967) and Neuts (1981), where a transform type analysis is carried out; a summary of results around DSP arrival processes can be found in Rolski (1989), where a stress on stochastic comparison was placed. Another approach was introduced by Livny et.al. (1993), for simulation purposes.

The present paper is directly related to Patuwo et al. (1993), and Szekli et al. (1994a,1994b), where the question of how dependencies in the arrival process influence the performance measures of MR/GI/1 queues was considered. We

* Department of Industrial Engineering, HANYANG UNIVERSITY

1) This paper was partially supported by Hanyang University.

introduce a special type of MR arrival processes, which is flexible enough to see how the lag-r correlation for the arrival process can be changed by different parameters, but which guarantees that the one dimensional distributions of the interarrival sequence remain the same after these changes.

2 Structure of the Model and Comparison Results

We consider the MR/GI/1 queue for a FCFS queue with a Markov renewal (MR) arrival process, which we define as follows. Consider a sequence of arrays of independent positive random variables $\mathcal{X} = \{[X_{ij}^{(k)}]_{i,j\geq 1},\ k = 1,2,\cdots\}$. For each i, j the sequence $\{X_{ij}^{(k)},\ k = 1,2,\cdots\}$ is i.i.d. with a common distribution function F_{ij}. Let $\mathcal{Z} = \{Z_0, Z_1, \cdots\}$ be a Markov chain with the state space $\{1,2,\cdots\}$, transition probabilities

$$a_{ij} = P(Z_k = j \mid Z_{k-1} = i),\ k \geq 1, \tag{1}$$

and an initial distribution $a_i = P(Z_0 = i)$. We assume that $\mathcal{Z}$ is irreducible, positive recurrent, and independent of $\mathcal{X}$. We denote by $\boldsymbol{\pi} = \{\pi_i,\ i = 1,2,\cdots\}$ the unique invariant probability measure for $\mathcal{Z}$.

Definition 2.1 *The arrival process with interpoint distances $D_k = X_{Z_{k-1},Z_k}^{(k)}$, $k = 1,2,\cdots$, is a Markov renewal (MR) arrival process.*

In a Markov renewal arrival process there are many types of arriving customers. Successive arrival types form a Markov chain and their interarrival times depend on the arrival types.

The distribution of the arrival process $\mathcal{D} = \{D_1, D_2, \ldots\}$ is uniquely determined by the initial distribution $\mathbf{a} = \{a_i\}$, the transition matrix $\mathbf{A} = \{a_{ij}\}$, and the set of distribution functions $\mathbf{F} = \{F_{ij}\}$. We denote this triple by $[\mathbf{a}, \mathbf{A}, \mathbf{F}]$. The corresponding semi-Markov kernel is $\mathbf{A}(t) = \{a_{ij}F_{ij}(t)\}$. We assume a special form of the transition matrix for the governing Markov chain

$$\mathbf{A}_n(p) = \begin{bmatrix} p & \frac{1-p}{n-1} & \cdots & \frac{1-p}{n-1} \\ \frac{1-p}{n-1} & p & \cdots & \frac{1-p}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1-p}{n-1} & \frac{1-p}{n-1} & \cdots & p \end{bmatrix}, \tag{2}$$

where $n \geq 2$, $p \in (0,1)$, and, for simplicity, F_{ij} depends only on j.
Consider a MR arrival process represented by $[\boldsymbol{\pi}, \mathbf{A}_n(p)6, \mathbf{F}]$ where $\boldsymbol{\pi}$ is the steady state distribution of the Markov chain (2). Then $\boldsymbol{\pi} = \frac{1}{n}\mathbf{e}$ and the marginal distribution for the stationary interarrival time, $P(D_k \leq t)$, $k = 1, 2, \ldots$ is given by

$$P(D_k \leq t) = \boldsymbol{\pi}\mathbf{A}(t)\mathbf{e} = \frac{1}{n}\sum_{i=1}^{n} F_i(t),$$

which is independent of p.
The parameter $p \in (\frac{1}{n}, 1)$ plays the main operational role in our study of effects caused by the correlations in $\mathcal{D}_p$ on the performance of MR queues. We adopt the above MR arrival process $[\boldsymbol{\pi}, \mathbf{A}_n(p)6, \mathbf{F}]$ because we can pull out the pure effect of correlation in teh arrival process on the queueby changing p as shown below.

In Szekli et al.(1994a) it is shown that the lag-r correlation coefficient of the arrival process is given by

$$corr_p(r) = \frac{\frac{1}{n^2}\sum_{i<j}(m_i - m_j)^2}{\frac{1}{n}\sum_{i=1}^{n} v_i + \frac{1}{n^2}\sum_{i<j}(m_i - m_j)^2}\left(\frac{np-1}{n-1}\right)^r, \tag{3}$$

where $m_i = \int_R x\, dF_i(x)$ and $v_i = \int_R (x - m_i)^2\, dF_i(x)$ for $i = 1, 2, \cdots, n$.
Thus we can increase correlations in $\mathcal{D}_p$ by increasing p while we keep the marginal interarrival distribution $P(D_k \leq t)$ the same.

In general the presence of positive correlations in $\mathcal{D}_p$ results in more variable waiting times. Szekli et al.(1994b) showed this by comparing queues: the MR arrival processes $[\boldsymbol{\pi}, \mathbf{A}_n(\frac{1}{n}), \mathbf{F}]$ ($\mathcal{D}_{\frac{1}{n}}$, i.i.d., no correlations), $[\boldsymbol{\pi}, \mathbf{A}_n(p), \mathbf{F}]$, $p \in (\frac{1}{n}, 1)$ (associated $\mathcal{D}_p$) and $[\boldsymbol{\pi}, \mathbf{A}_n(1), \mathbf{F}]$.

For a stationary ergodic MR/GI/1 queue with the MR arrival process given by $[\boldsymbol{\pi}, \mathbf{A}_n(p), \mathbf{F}]$, i.i.d sequence of service times $\{S_n\}$, and $ES_1/ED_1 < 1$, denote by W_p the stationary actual waiting time in this queue. Note that the traffic intensity $\rho = ES_1/ED_1$ does not depend on p in this class of queues.

Theorem 2.2 *(Szekli et al.(1994b)) Consider three stationary ergodic MR/GI/1 queues ($\rho < 1$) with the interarrival processes represented by* $[\boldsymbol{\pi}, \mathbf{A}_n(\frac{1}{n}), \mathbf{F}]$, $[\boldsymbol{\pi}, \mathbf{A}_n(p), \mathbf{F}]$ *($p \in (\frac{1}{n}, 1)$), and* $[\boldsymbol{\pi}, \mathbf{A}_n(1), \mathbf{F}]$, *respectively. Then*

$$EW_{\frac{1}{n}} \leq EW_p \leq EW_1.$$

3 Role of p and $m_i - m_j$

We restrict now our attention to MR/M/1 queues with the arrival stream $[\pi, \mathbf{A}_2(p), \mathbf{F}]$. This two dimensional setting allows us to find a closed form for expectation of the stationary queue length embedded at the arrival times.

Let the semi-Markov kernel be

$$\mathbf{A}(t) = \begin{bmatrix} pF_1(t) & (1-p)F_2(t) \\ (1-p)F_1(t) & pF_2(t) \end{bmatrix},$$

where $F_i(t) = 1 - e^{-\lambda_i t}$, $i = 1, 2$, $t \geq 0$. Suppose that the service time is exponentially distributed with the rate μ. The "individual" traffic intensities are $\rho_1 = \lambda_1/\mu$, and $\rho_2 = \lambda_2/\mu$.

For this setting, we have two types of arrival customers. The successive arrival types form a Markov chain

$$\mathbf{A}(t) = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix},$$

and the interarrival time distribution when the arriving customer is of type i, $i = 1, 2$, is F_i.

The traffic intensity for the system is the harmonic mean of the individual traffic intensities $\rho = \{\frac{1}{2}(\frac{1}{\rho_1} + \frac{1}{\rho_2})\}^{-1}$. We assume that $\rho < 1$. Then it is shown in Szekli et al.(1994b) that L^t, the mean queue length at arbitrary times in the steady state, becomes

$$L^t = \frac{\rho}{1-\rho} + \frac{\rho}{2(1-p)(1-\rho)}\left(1 - \frac{2(1-P_0)}{\rho_1 + \rho_2}\right). \tag{4}$$

From the above result we obtain the following important theorem, whose proof can be found in Szekli et al.(1994b)

Theorem 3.1 *Consider a stationary MR/M/1 queue with the arrival process* $[\pi, \mathbf{A}_2(p), \mathbf{F}]$ *and* $\rho < 1$. *If the arithmetic mean* $(\rho_1 + \rho_2)/2 > 1$ *then*

$$L^t \to \infty \text{ as } p \to 1.$$

Remark. The first term $\rho/(1-\rho)$ in the formulas in (4) is the mean stationary queue length of M/M/1 queue. Thus, we see the second term is due to the dependency in the arrival stream.

Theorem 3.1 shows the queue length can be infinitely large even if $\rho < 1$ in a correlated arrival case.

Now we explain the queueing behavior in terms of $m_2 - m_1$ and p by the following corollaries of Theorem 3.1.

Corollary 3.2 *If the difference of two means m_1 and m_2 is great enough, then $L^t \to \infty$ as $p \to 1$.*

Proof. Without loss of generality, we hereafter assume $m_2 - m_1 > 0$. Since $\mathrm{E}D_n = m = \frac{1}{2}(m_1 + m_2)$ and $\rho = 1/(\mu \mathrm{E}D_n)$, we have $m_1 + m_2 = \frac{2}{\mu\rho}$.
Now, the condition $(\rho_1 + \rho_2)/2 > 1$ in Theorem 3.1 is equivalent to $\frac{1}{m_1} + \frac{1}{m_2} > 2\mu$ and in turn,

$$m_1 m_2 < \frac{m_1 + m_2}{2\mu} = \frac{1}{2\mu}\frac{2}{\mu\rho} = \frac{1}{\mu^2\rho}. \tag{5}$$

Then $m_2 = \frac{1}{2}\left(\frac{2}{\mu\rho} + d\right)$ and $m_1 = \frac{1}{2}\left(\frac{2}{\mu\rho} - d\right)$. So (5) becomes

$$\frac{1}{4}\left(\left(\frac{2}{\mu\rho}\right)^2 - d^2\right) < \frac{1}{\mu^2\rho}.$$

Rearranging both sides yields $d^2 > \frac{4}{\mu^2\rho}\left(\frac{1}{\rho} - 1\right)$. Since $d > 0$,

$$d > \frac{2}{\mu}\sqrt{\frac{1}{\rho}\left(\frac{1}{\rho} - 1\right)} = 2m\rho\sqrt{\frac{1}{\rho}\left(\frac{1}{\rho} - 1\right)} = 2m\sqrt{1-\rho} \equiv c. \tag{6}$$

$m_2 - m_1 = d > c$ is equivalent to $(\rho_1 + \rho_2)/2 > 1$, which completes the proof. □

Corollary 3.3 *When the difference $m_2 - m_1$ is small enough, the queue is stable.*

Proof. Assuming $m_2 > m_1$, we have

$$0 < m_1 < m < m_2 < 2m. \tag{7}$$

There is no chance for the mean queue length to be infinitely large if $\rho_1 < 1$ and $\rho_2 < 1$, that is,

$$m_1 > m\rho, \quad m_2 > m\rho. \tag{8}$$

Combining (8) with (7) we have $m\rho < m_1 < m < m_2 < 2m$, which, in terms of $m_2 - m_1$, becomes $0 < m_2 - m_1 < (2m - m\rho) - m\rho = 2m(1-\rho) \equiv c'$. □

We give some numerical examples to make the above interpretation clear.

Example 1. In this example we illustrate the joint effect of $m_2 - m_1$ and p and see how the critical and safety regions move as the traffic intensity varies.

We fix the mean value m of the marginal interarrival times to be 5. Recall that $m = 1/2(m_1 + m_2)$. Since $m_1 + m_2 = 10$ and we assume $m_2 > m_1$, $5 < m_2 < 10$ and $0 < m_1 < 5$.
Then $c = 2m\sqrt{1-\rho} = 10\sqrt{1-\rho}$ and $c' = 2m(1-\rho) = 10(1-\rho)$.

For the traffic intensities $\rho = 0.1,\ 0.5,\ 0.9$, we give different values of $m_2 - m_1$ and compute L^t, the mean queue length at arbitrary times. Figures 1 to 3 show the behavior of L^t versus p which varies from 0.5 to 1.0 (We plot only the range of p from 0.75 to 1.0 to make the graph more clear).

For $\rho = 0.1$, $c = 9.49$ and $c' = 9$. Thus under light traffic intensity, the queue stays stable for most of the possible m_1 and m_2 values (see Figure 1).

For moderate traffic intensity $\rho = 0.5$, $c = 7.07$ and $c' = 5$. (see Figure 2).

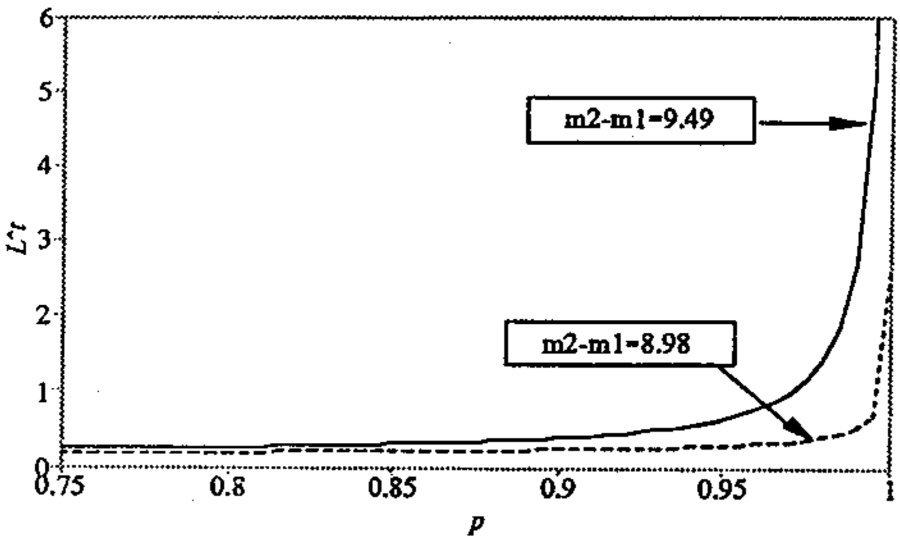

Figure 1: The behavior of L^t vs. p ($\rho = 0.1$)

Now under a heavy traffic intensity $\rho = 0.9$, $c = 3.16$ and $c' = 1$. (Figure 3). Thus, we conclude that under heavy traffic, we have more chance for the queue length to be arbitrarily large, while under light traffic, the queue tends to be stable.

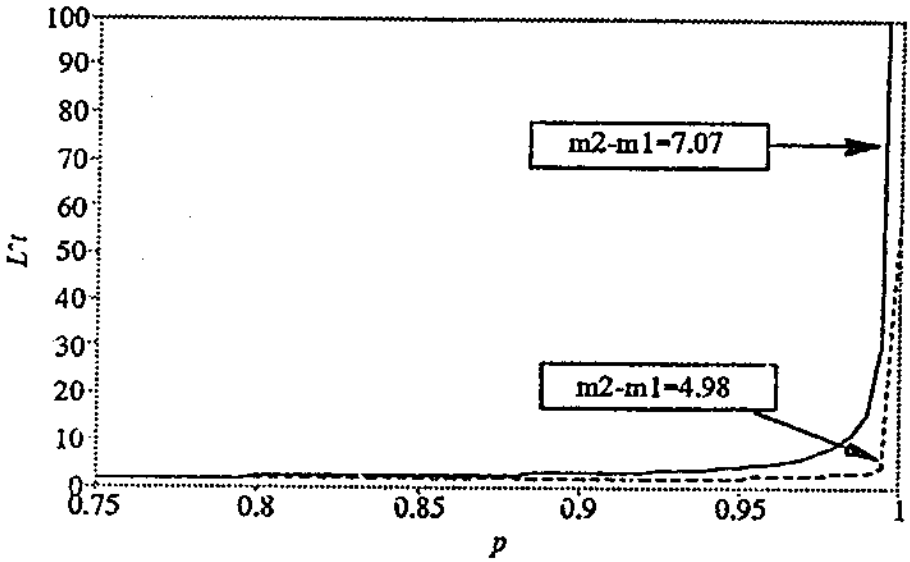

Figure 2: The behavior of L^t vs. p ($\rho = 0.5$)

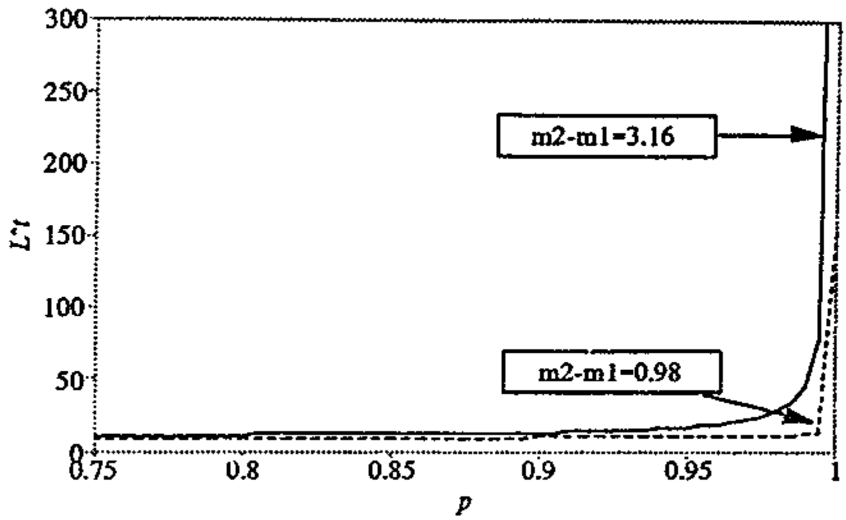

Figure 3: The behavior of L^t vs. p ($\rho = 0.9$)

4 Change of Variability

In the previous Section 3, we showed that the increase of the correlation coefficient in the arrival process via the parameter p, jointly with $m_2 - m_1$, the distance between mean values, can make the queue quite unstable. However, we may increase the correlation coefficient by reducing the variances of the F_j's, which gives a different result. The following lemma is a direct application of Rolski(1983) to our $\mathbf{A}_n(p)$.

Lemma 4.1 *Consider two MRAPs* $[\boldsymbol{\pi}, \mathbf{A}_n(p), \mathbf{F}]$ *and* $[\boldsymbol{\pi}, \mathbf{A}_n(p), \mathbf{F}']$ *to a single server queueing system, where* $F_j <_{icx} F'_j$ *and the expected value of* F_j *and* F'_j *are equal for* $j = 1, \ldots, n$. *Then for the corresponding correlation coefficients and stationary actual waiting times we have*

$$\mathrm{Corr}[\boldsymbol{\pi}, \mathbf{A}, \mathbf{F}] \geq \mathrm{Corr}[\boldsymbol{\pi}, \mathbf{A}, \mathbf{F}'],$$

while

$$\mathrm{E}W[\boldsymbol{\pi}, \mathbf{A}, \mathbf{F}] \leq \mathrm{E}W[\boldsymbol{\pi}, \mathbf{A}, \mathbf{F}'].$$

Proof. The second part of the lemma is immediate from Rolski(1983).

The variance of F_j is smaller than that of F'_j by assumption $F_j <_{icx} F_{j'}$. Recalling that the formula for correlation coefficient (3), since each correlation coefficient has the sum of variances of F_j's, $\sum_{i=1}^{n} v_i$, in its denominator, the first part of the lemma follows. □

The Lemma 4.1 tells us that a larger correlation coefficient in the arrival process does not necessarily cause a larger mean queue length. This happens when we change the correlation coefficient by changing the variances of the interarrival times, which is analogous to the Pollaczek-Khintchine formula.

Example 2. In this example, noting that the variance of $\mathrm{Er}(k, \lambda)$ decreases as k increases when we keep λ fixed, we illustrate Lemma 4.1 that the correlation coefficient can increase while the mean queue length decreases.

Consider a two-state ($n = 2$) MRAP of which the transition matrix of the underlying Markov chain is given by:

$$\mathbf{A} = \begin{pmatrix} 0.85 & 0.15 \\ 0.15 & 0.85 \end{pmatrix},$$

i.e., $p = 0.85$. Let $F_1 = \text{Er}(k, 1/3)$ and $F_2 = \text{Er}(k, 1/7)$ and the traffic intensity is $\rho = 0.5$. Then the lag-1 correlation coefficient is

$$\text{Corr} = \frac{(7-3)^2}{\frac{2}{k}(7^2+3^2)+(7-3)^2}\left(\frac{2\cdot 0.85-1}{2-1}\right) = \frac{11.2k}{16k+116},$$

which is increasing and concave in k. Since for $k < k'$, we have $\text{Er}(k', 1/3) <_{icx} \text{Er}(k, 1/3)$ and $\text{Er}(k', 1/7) <_{icx} \text{Er}(k, 1/7)$, $\text{E}W(k) \leq \text{E}W(k')$ by Rolski(1983). By Little's result, $L^t(k) \leq L^t(k')$, that is, L^t decreases in k, as Figure 4 illustrates.

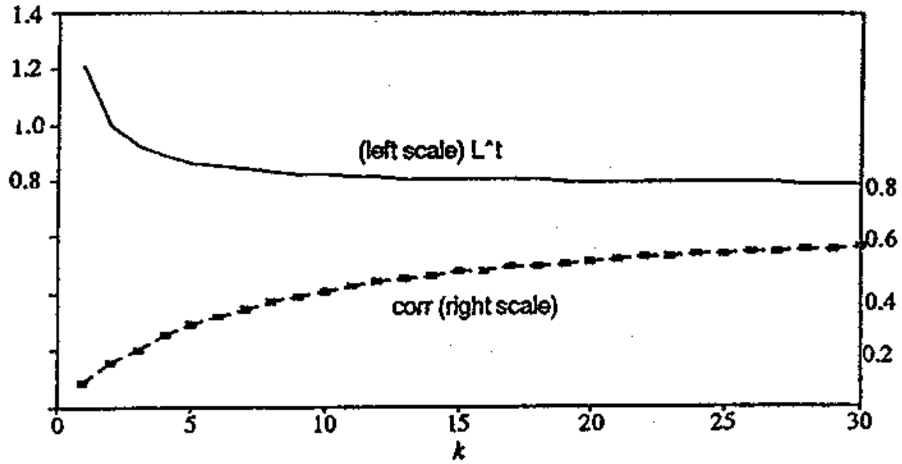

Figure 4: The behavior of L^t and Corr vs. k

5 Summary and Results

In this research we have considered correlated arrival queueing systems to see the effect of this correlation on the queueing performance. We introduced a special structure of a Markov renewal arrival process whose transition matrix of underlying Markov chain is $\mathbf{A}_n(p)$. This guarantees the marginal interarrival time distributions are the same.

The effect of p, the jump intensity from one type to another, becomes dramatic when combined with the magnitude of $m_i - m_j$, the differences of the means of the interarrival times of different types. Even though the solution is case specific, it is enough to draw important interpretation. When the difference $m_2 - m_1$ is large enough and p increases to one (or, the correlation coefficient grows through p) the queue length becomes arbitrarily large even though the overall traffic intensity is smaller than one (so the queue is said to be stable). This is because once the arriving customer is of type 1 (a heavy traffic state) then the successive arrivals tend to be of the same type, and the heavy traffic state during the arrival of type 1 customers dominates the queue length. The low traffic state during the arrival of type 2 customers cannot return the queue to be stable. We conclude from this observation that the high correlation coefficient via high p may yield a very significant effect. This result allows us to construct examples for which the mean queue length of the correlated arrivals can be arbitrarily many times larger than the corresponding uncorrelated arrival queue.

A larger correlation coefficient, however, does not necessarily imply larger queue length, as we have shown in Section 4 through our $\mathbf{A}_n(p)$ structure. The larger correlation coefficient obtained by making the variance v_i, the variance of interarrival times for each type smaller causes a smaller queue length. If we interpret the variance as a measure of uncertainty, then our result says that by reducing the uncertainty in the arrival stream we can obtain a smaller queue length.

References

Çinlar, E. (1967) Queues with semi-Markovian arrivals. *J. Appl. Prob.* 4, 365-379.

Disney, R.L. and Kiessler, P.C. (1987) *Traffic Processes in Queueing Networks, A Markov Renewal Approach.* The Johns Hopkins University Press. Baltimore.

Livny, M., Melamed, B. and Tsiolis, A.K. (1993) The impact of autocorrelation on queueing systems. *Management Science* **39**, 322-339.

Meester, L. and Shanthikumar, J.G. (1993) Regularity of stochastic processes: a theory based on directional convexity. *Prob. in the Eng. and Informational Sciences* **7**, 343-360.

Neuts, M.F. (1981) *Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach.* The Johns Hopkins University Press. Baltimore.

Patuwo, B.E., Disney, R.L., and McNickle, D.C. (1993) The effect of correlated arrivals on queues, *IIE Transactions* **25**, 105-110.

Rolski, T. (1983) Comparison theorems for queues with dependent interarrival times, in: *Modelling and Performance Evaluation, Proceedings of the International Seminar*, Paris, 42-67.

Rolski, T. (1989) Queues with nonstationary arrivals. *Queueing Systems* **5**, 113-130.

J.G. Shanthikumar and D. Yao, (1991) Strong stochastic convexity: closure properties and applications, *J. Appl. Prob.* **28**, 131-145.

Stoyan, D. (1983) *Comparison Methods for Queues and Other Stochastic Models.* John Wiley, Berlin.

Szekli, R., R. L. Disney and S. Hur. (1994a) On performance comparison of MR/GI/1 queues. *Queueing Systems:Theory and Applications*, **17**, 451-470.

Szekli, R., R. L. Disney and S. Hur. (1994b) MR/GI/1 queues with positively correlated arrival streams. *J. Appl. Prob.* **31**, 497-514.

95년 8월 최초 접수, 95년 11월 최종 수정